Summary of project

• Conduct 3 part 1 exams using online computers
• The division exams are labeled X, Y and Z in this presentation.
• Some used single center delivery and some multiple centers
• Sites in UK, Greece, Hungary, Poland and India
• Some used bring your own computers and some existing computer labs
Contract

• Orzone
• Provide software
• Provide training
• No manual labor
• Do part 1 and part 2 online
• Do all divisions

• UEMS
• Should invest time to train staff on use of Orzone tools and exam methods
• Have one point of contact for all divisions
• All similar steps done at the same time
Summary findings

• Positive
• All data was successfully captured
• Sites in Europe and Asia
• One of 3 exam had good content quality
• Passionate people

• Need to improve
• Many critical processes are ad-hoc
• Item quality need to improve for 2 of 3 exams
• Workflow for management of administrative process.
• Workflow to achieve good quality in writing and reviewing
• Cost-effectiveness. UEMS Surg is about 5-10 times more labor intensive per exam than other European boards.
• Attention to details
• Risk management
Before, during and after exam

- Prepare Venue
- List of candidates
- List of questions

- Statistics
- Pass Fail
- Report and diploma
Key processes prior to exam
Questions/Items

- Test design
- Organization
- Write
- Review
- Select final questions
- Mock exam
Venue

• One or more physical sites
  • More centers make it easier for candidate to sit exam = growth
  • Tables, power adapters

• Invigilator
  • Check ID and supervise
  • Assist in set-up of room

• IT
  • Wifi and optional back-up wifi
Candidate processing

• Registration/Payment

• Vetting/eligibility/portfolio

• Information
  • Visa letter for non-EU
  • Hotel/travel
  • Venue info
  • Web, social media and email
  • How to use computer
  • Blueprint of exam
Success factors during exam

• Multiple centers
• Separate part 1 and part 2
• Share part 1 dates to lower cost
• Non-EU centers
• Back-up wifi
• Lock-down browser
Key processes after exam
How to measure quality of an exam

• Analyze whole paper
  • Reliability, validity
  • Example: KR20, Cronbach alpha, Sum of measurement error.

• Analyze individual questions/items
  • Item discrimination: Top/Bottom or Pearson-r
  • Distractor discrimination: histograms
  • Difficulty: % that gets it right

• Other statistics
  • Histograms, mean, standard dev
  • Etc Etc
Example quality measure of benchmark exam

- Average item discrimination high (e.g. each question statistically separate strong from poor candidates)
- Few questions with negative discrimination, (negative item discrimination means the questions statistically selects poor candidates)
- Lesson is that you don’t need to write perfect questions, but on average good and have few poor questions
- A good exam is a trade-off between nr of questions and the quality of the questions

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average item discrimination</td>
<td>0,15</td>
<td>0,09</td>
<td>0,12</td>
</tr>
<tr>
<td>Ratio poor items</td>
<td>6%</td>
<td>17%</td>
<td>23%</td>
</tr>
</tbody>
</table>

A is a leading European board Orzone use as a benchmark of excellence B and C are two smaller boards that are more new to exams. Orzone use these to get a reference interval of acceptable levels to also take into account less resources
UEMS Surg vs benchmark

- Division X and Y are a factor of 4-5 lower in average item discrimination and a factor.
- X and Y have 6-8 times as many items performing poorly. The ratio is so high that the exam should be scrutinized.
- Common examples of items performing poorly are NOT questions, negations, False questions, too easy/difficult.
- The 3 exams used different methods and extent of review and statistics.
Other measures

Volume of emails. External indicator on internal processes

<table>
<thead>
<tr>
<th></th>
<th>Board A</th>
<th>Board B</th>
<th>Board C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr of emails per candidate</td>
<td>0.02</td>
<td>0.39</td>
<td>0.67</td>
</tr>
</tbody>
</table>

A, B And C above also include different registrations, Portfolios, payments and printed diplomas.
Comparison with UEMS Surgery

Volume of emails. External indicator on internal processes
Illustrate with a small portion of the process
Let us illustrate with a small portion of the process
Illustration of workflow of an UEMS section

IMPROVING THE HEALTH OF WOMEN AND THEIR BABIES

Part 1 exam (KBA)

May 17, 2018

The exam assess knowledge (KBA) and is computer based.

The exam can be done in Brussels, Belgium.

Passing this exam in necessary to sit the part 2 exam.

The number of seats are limited. Early reservation recommended.

Click here to register

No acronyms
From here on the candidate will now have to visit multiple sites, download up to 10 documents in 3 formats. This will be mail and/or clicked up to 10 times internally and externally. After that an excel sheet will be created manually that contain information about each candidate.

This information was in the end incomplete or altered for each of the 3 exams. Candidates was registered in wrong countries etc. In one case the candidates was given the wrong time on web page and different time in the invitation letter.
High quality correlate to high volume
High price correlate to high volume

- Implicitly because quality of exam is strongly linked to reputation which is important for volume
- Higher price also means more money to market exams
- Low price of exams is potentially spiral of death, e.g. sufficient resources can not be invested in quality and thus growth become impossible
Considerations on quality and risk

• Relationship between
  • Time, Price and quality

• Risk management vs quality
  • What is someone sues and process is not in order.
  • 10 candidates incorrectly processed could means several hundred thousand Euros in damages.
Main recommendation

• Improve exam quality
  • **Recruit experienced staff** to assist doctors

• Improve admin process (e.g. less errors and less work)
  • **Complete reworking of workflow**

• Focus on the quality of the whole exam
  • Set the price after what it actually cost to deliver quality

• Improve exam quality
  • **Recruit experienced staff** to assist doctors to achieve high quality questions and offload doctors with statistics, item analysis etc.
  • 5-10 days need to be invested in each division exam (e.g. half to full-time)

• Improve admin process (e.g. less errors and less work)
  • **Complete reworking of workflow** for how the different parts of the exam process is managed

• Focus on the quality of the whole exam and not the bits and pieces
  • Nasa video
  • The first recommendation above will increase cost and quality
  • The second recommendation above will reduce cost and increase quality
  • Set the price after what it actually cost to deliver quality
  • Quality lead to volume
Detailed Recommendations

**Improve non-medical administration of exam**

- Create new simplified workflow. See image of current workflow.
- Either train current staff or recruit additional staff with experience from managing medical educational events/exams.
- First fix quality. Then plan for growth. Growth necessary for international recognition and to provide funding to develop better exams in the future.
- Integrate part 1 and part 2. Rollout plan.
- Harmonize dates at multiple centers (e.g. run part 1 on same dates) for exams to reduce cost and administration.

**Improve medical content**

- Ensure that doctors have appropriate support. Doctors shall focus on content, not admin.
- Recruit lead exam content manager. At least 5 years experience from medical board exams and psychometrics.
- Establish new workflows for writing and review.
- Make surveys at end of exam to get input from candidates
- Provide training in how to write good questions.
- Reduce complexity by harmonizing exam format across divisions. Make it easier and lower cost to deliver.
- Involve exam supplier in exam meetings prior to exams.
Root cause of **success**

Activities that drive
- Awareness
- Solve logistical bottlenecks
- Make it easy for candidate to sit exam
- What are common factors contributing to high number of candidates and growth

Ensuring that candidates know there is an exam
- Clear growth plan with concrete marketing activities. Web, Facebook and Twitter. Special exam for in training assessment to make residents used to exams.

Ensure that it is easy for a candidate to sign up
- Automated registration and eligibility process. No emails.
- Exams are run at multiple centers. Test centers outside EU. Exams run multiple times per year. Closer for candidate.

Make it easier to deliver exams
- All exams delivered online, BYOD.
- Separation of part 1 and part 2.

Improved internal exam processes and other factors
- Involve exam supplier as exam advisor in board planning
- Exam price higher than EU average to cover expenses for professional delivery.
- Special exam for in training assessment to make residents used to exams.
Root cause of quality

Quality means that the questions used in the exam are of good quality

Quality also means that there are robust processes in place

Quality also means that there is sufficiently trained staff.

The number of emails sent is an indicator of how well the processes are running.

• What are common factors contributing to high number of candidates and growth
  • All exams delivered online, BYOD.
  • Exams are run at multiple centers.
  • Test centers outside EU.
  • International strategy for translations.
  • Exam price higher than EU average to cover expenses for professional delivery.
  • Separation of part 1 and part 2.
  • Exams run multiple times per year
  • Special exam for in training assessment to make residents used to exams.
  • Automated registration and eligibility process
  • Clear growth plan with concrete marketing activities
Empirical success factors

- How can exam success be measured
  - Nr of exam candidates
  - Nr of exam candidates vs available European doctors (e.g. penetration)
  - Growth of candidates
- How can exam quality be measured
  - Item discrimination – ratio of weak questions
  - Item discrimination – ratio of good questions
  - Nr of emails sent per candidate – measure of functioning workflow.

Root cause vs factor